

# Mathematical formulation of the purchase funnel by using knowledge space theory

Yung-Tzu Lin, Chuan-Yi Chang, Shein-Yung Cheng, Meng-Yun Lin

Lin YT, Chang CY, Cheng SY et al. Mathematical formulation of the purchase funnel by using knowledge space theory. *J Pure Appl Math.* 2023; 7(1):01-19.

## ABSTRACT

The retail industry provides customers with goods or services. Analyzing the purchase behavior of customers is critical for expanding the business. Therefore, managing the repurchase intention of customers is crucial. A sequence of purchase behaviors by each customer constitutes a set of purchase Customer Journeys (purchase CJs), which detail purchase pathways and repurchase behaviors. Purchase CJs are the actual retail transaction data. This study investigated purchase CJs and proposed a purchase funnel called CJ graph (CJG) by measure theory and knowledge space theory with actual retail transaction data. To achieve this objective, a Customer Journey Block (CJB), denoting "touchpoint," is defined as a series of purchase behaviors during a period and used as the base of this method. Each touchpoint is allotted a measurable function, named Purchase Measure (PM). By integrating all the CJBs of purchase CJs, the Purchase Measure Graph (PMG) can be constructed as the primary structure of the PM knowledge structure. Finally, when all CJs are coordinatized with CJ codes, knowledge space theory is used to develop the Customer Journey Graph (CJG). In this method, knowledge bases are used as the spanning framework of the CJ knowledge space,

and the filtering property of the purchase funnel is illustrated. Furthermore, to validate the feasibility of the proposed method, a set of two-year and one million transactions retail data generated from more than thirty-six thousand customers was segmented by Recency, Frequency, and Monetary (RFM) model. Thus, the corresponding PMGs and CJGs of various (12, here) customer segments are detailed, and the resultant knowledge bases of the primary purchases and secondary re-purchases are applied to analyze and predict the customer purchase behaviors accurately.

**Keywords:** *Transaction; Purchase measure; Customer journey; Knowledge space theory*

## Abbreviations

AIDA	Awareness, Interest, Desire, and Action
CJ	Customer Journey
CJB	Customer Journey Block
CJG	Customer Journey Graph
EC	Equivalence Class
KST	Knowledge Space Theory
NB	National Brand
PM	Purchase Measure
PMG	Purchase Measure Graph
PMKS	Purchase Measure Knowledge Space
RFM	Recency, Frequency, and Monetary
VIP	Very Important Person

## INTRODUCTION

Retail caters to making products available for customers to purchase. However, in a conventional shop floor, when a customer completes the transaction and leaves the store or website, the seller can only hope the customer visits again. Therefore, investigating the customer's repurchase behavior and constructing a model of multiple transactions of customers is critical. In retail, customers buying a certain product is a behavioral tendency of their purchase intention [1]. Repurchase intention or repeat patronage indicates the possibility that customers will repurchase products or

avail services of the same brand again [2].

## Purchase and repurchase in retailing

### Repurchase characteristics

Although repurchase intentions and purchase intentions may appear to be similar consumer behaviors, they have distinct connotations. The purchase intention is a subjective tendency of consumers to consider the purchase behavior, which can be used as a key indicator to predict consumer purchase behavior [3,4]. Repurchase intention is the possibility of customers maintaining the frequency of past

Department of Information and Computer Engineering, Chung Yuan Christian University, Taoyuan City 320314, Taiwan (R.O.C.)

Correspondence: Chuan-Yi Chang, Department of Information and Computer Engineering, Chung Yuan Christian University, Taoyuan City 320314, Taiwan (R.O.C.), e-mail chang.maverick@gmail.com

Received: Jan 16, 2023, Manuscript No. puljpm-23-6058, Editor Assigned: Jan 18, 2023, PreQC No puljpm-23-6058(PQ), Reviewed: Jan 19, 2023, QC No. puljpm-23-6058 (Q), Revised: Jan 20, 2023, Manuscript No puljpm-23-6058 (R), Published: Jan 30, 2023, DOI:10.37532/2752-8081.23.7(1).01-19.



This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to noncommercial purposes. For commercial reuse, contact [reprints@pulsus.com](mailto:reprints@pulsus.com)

consumption and continuing to purchase the same product [5]. In addition to repurchase intention, the customer simultaneously exhibits other behaviors such as word of mouth and public recommendation [6].

#### Definition of repurchase

When the customer is satisfied with the product or service, then repeated purchase behavior, that is, repurchase intention may occur [7]. Researchers have revealed that customer satisfaction considerably affects purchase intention. High satisfaction positively affects repurchase willingness and can stabilize customers' repeated purchase (repurchase) behavior [8, 9]. The sense of identity is another critical factor. When the customer's sense of identity is higher, they are willing to continue to purchase more products at higher prices. Therefore, a high sense of identity positively influences the willingness to increase prices [10].

Researchers believe that repurchase intentions are “derivative behaviors of loyalty,” which denotes willingness to continue buying and recommending products to others. Therefore, the positive definition of repurchase intention is the derivative behavior of customer loyalty, that is, consumers publicly introduce or recommend to others [11]. For a company, customer loyalty is a critical indicator. The parameter indicates that customers will continue to patronize, repeatedly, or exclusively purchase and use the company's products and services, and introduce the product to their friends [12].

#### Repurchase model

Increasing customer willingness to buy back is crucial for companies because the finding new customers is approximately five times expensive than retaining existing customers [13,14]. For saving corporate costs, reducing the loss of customers is helpful to the company's profit than reducing costs [15]. The repurchase intentions of customers can be measured from three aspects, namely:

1. Intents to repurchase: This behavior measures customers' willingness to repurchase the company's products or services in the future, which is an indicator of customers' future behavioral intentions.
2. Primary behaviors: This behavior includes the number of purchases, frequency, amount, and quantity of customers.
3. Secondary behaviors: This behavior refers to behaviors that customers help the company introduce, recommend, and build reputation. Such behaviors are valuable to the company [16].

Among the three behaviors, only primary behaviors can be measured by actual transaction data. For each customer, his/her primary behaviors change over time. Thus, each customer has a distinct CJ.

### **CJ**

#### CJ characteristics

In general retail purchases, customer experience includes not only the purchase of goods but also the customer's cognitive, affective, emotional, social, and various physical responses to the purchased brand (service provider) [17]. Furthermore, CJ collects the procedural and empirical aspects from the customer perspectives of these products or services provided by the brand [18]. Formally, this CJ can be described as repeated interactions between customers and the brand (service provider) [19]. Thus, the entire CJ becomes an “engaging story” about the user's interaction with the brand or can be

termed as walking “in the customer's shoes.” [20, 21].

#### Definition of CJ

The interactions or communications between customers and a brand (service provider) can be defined as “touchpoints” which represents the abstract form of customer experience (the customer's product purchases in this paper) [22-24]. These touchpoints form the building blocks of CJ, which is then described as a series of touchpoints [25-28]. When combined with data analysis, CJ can be visualized as a CJ map [19-20]. The CJ map can be visualized in a stream, in which the touchpoint represents the abstract form of the customer experience (the customer's product purchases in this paper) [25].

#### CJ of repurchase

From the perspective of customer purchase analysis, CJ is an open process comparable with the “customer loyalty staircase” [29]. In retail analytics, the evaluation of customer purchase journeys is essential CJ is a clearly defined service process, with marked starting and ending points [30-31]. In the CJ of purchase analysis, the initial touchpoint is the first purchase behavior of a new customer, and the final touchpoint is the purchasing characteristic of loyal customers. Evaluating these topics is essential for researching purchase funnels.

#### **Purchase funnel**

##### Marketing funnel

The relationship between the funnel model and the Awareness, Interest, Desire, and Action (AIDA) was first proposed in the early twentieth century [32]. Marketing funnel theory originates from the hierarchy of effects-based marketing theory, which categorizes the mental state of the consumer into various marketing procedures: from brand awareness, consideration of purchase, to brand preference, and purchase action to achieve brand loyalty [33]. Specifically, consumers performing purchase decisions can be categorized into an AIDA process:

1. Awareness: the customer is aware of the existence of a product or service,
2. Interest: actively express interest in a certain product group,
3. Desire: desire for a specific brand or product,
4. Action: product purchase action [34].

Therefore, using the effect hierarchy of AIDA, the marketing funnel, based on the process of marketing activities and psychological evolution of consumer purchase activities, is used to specifically construct a repeatable, traceable, highly operable customer purchase journey model to improve marketing efficiency.

##### Conversion funnel

Similar to a marketing funnel, in e-commerce, a “conversion funnel” is used to describe the trajectory of consumers browsing e-commerce websites through Internet advertising or search systems, and finally converting the search into sales [35]. In digital marketing funnel, consumers may find products they desire through advertisements or keyword searches on the web. Subsequently, social media is used to compare prices and consumer reviews to target websites or platforms. However, the purchase may occur in physical stores or online. This shopping experience and after-sales customer service determines if the customer will return. In practice, the marketing funnel, which is a conversion funnel, is a

visual interactive tool that represents the key conversion process from sales leads to sale completion. Figure 1 details the conversion ratio of e-commerce stores from online impressions (92,000 people) to actual orders received on the website (42 people). The conversion in each of the two stages provides insights into sales performance. For example, the pay-per-click marketing activities is related to the number of nonrepetitive visitors from online impressions (92,000 people) to the website (more or less than 3,200 people). This process is not limited for only e-commerce. Similar conversion funnels can be used for business-to-business companies or physical retail stores [36].



Figure 1) Conversion funnel of e-commerce

The “purchase funnel,” “sales funnel,” “marketing funnel,” or “customer funnel” is a consumer-focused marketing model that is used illustrate the theoretical CJ of customers purchasing products or services [35]. Studies have investigated multiple stages along the purchase funnel to investigate manufacturers perspective on managing National Brands (NBs) and the influence of in-store retailer-specific activities on the purchase decision [37]. Through purchase funnel, data analytics provide insights on customers, such as their behaviors, their incoming channels, and purchased contents [30].

Research objectives

Paper targets

Based on retail transaction data, this study proposed PM theory to construct the theoretical foundation of CJs, developed PM graph and customer journey graph (purchase funnel), and applied actual corporate data applications.

Contributions and objectives

The major contributions of this paper are as follows:

- (1) To propose algorithms for the formulation of the Purchase Measure Graph (PMG) and coordinate system to represent customers’ purchase behaviors.
- (2) To introduce pseudo-source design to integrate customer journeys into one connected graph and ensure PMG exhibits well-behaved properties.
- (3) To define the measurement functions, such as purchase frequency measure  $\mu F(.)$  and/or purchase monetary (the amounts of products) measure  $\mu M(.)$  of customer journey block

- (CJB), and the equivalence class, which is the purchase measure quotient space for formulating Purchase Measure Knowledge Space (PMKS).
- (4) To add quantitative values into the knowledge states of the knowledge space, such as coordinates, volumes, and customer flows of state transitions.
- (5) To provide an algorithm for the formal description of Customer Journey Graph (CJG) and to model purchase funnel by using a CJG.
- (6) To apply a two-year retail data from over thirty-six thousand customers to prove the feasibility of the purchase funnel and help the marketing strategy of retail business through the study of customer’s purchase journey by formulating the model of “marketing funnel” using the formal method.

Outlines

Section 2 formulates the purchase measures and forms their purchase measure space, Section 3 proposes a Purchase Measure Graph (PMG). By slightly modifying PMGs, Section 3 also introduces the knowledge space theory to develop the purchase measure knowledge space and analyze the corresponding knowledge base. For the knowledge states in the PM knowledge space, Section 4 presents a study of their quantitative values, including the coordinates, volumes, and flows of state transitions. Subsequently, the Customer Journey Graph (CJG or purchase funnel) is developed to visualize its filtering property. To prove the feasibility of the purchase funnel, a two-year retail data from over 36000 customers is used in Section 5 to establish PMGs and CJGs for the exploration of CJs of various customer segments. Section 6 presents the conclusion.

PM SPACE

Customer PM space

Transaction space

Let  $C = \{c_i\}$  be the set of customers,  $P = \{p_j\}$  be the set of products,  $M = \{m_j : m_j \in [0, M_{max}]\}$  be the payment, and the time  $t \in T$ , which denotes a time interval. A customer purchase behavior is defined as a quadruple as follows:

$$x = (c_i, \{p_j\}, \{m_j\}, t) \tag{1}$$

which indicates the customer  $C = \{c_i\}$  purchases  $\{p_j\} \subseteq P$  with the expense  $\{m_j\}$  at time  $t \in T$ . A transaction space  $\Xi(C, P, M, T) = C \times 2^P \times 2^M \times T$  encircles all the possible transactions  $x$ 's in Equation (1), in which  $2^P$  and  $2^M$  denote the power set of  $P$  and  $M$ , respectively. The following analysis begins with a transaction set  $X$  contained in the transaction space  $\Xi$ .

Example 1

When the transaction space is formed from the customer set  $C_A = \{c_1, c_2, c_3, c_4, c_5\}$ , the product category set  $P_A = \{g_1, g_2, g_3, g_4\}$ , the payment range  $M_{maxA}=499999$ , and the time interval  $T_A = [2019/01/01, 2020/12/31]$ . Here,  $g_1 - g_4$  are product categories, and not only products. The illustrated transaction space in Example 1 are presented in Table 1.

**TABLE 1**  
**Conversion funnel of e-commerce**

Customer	Date	Category	Amount
c <sub>1</sub>	1/16/2019	g <sub>1</sub>	1990
c <sub>2</sub>	7/6/2020	{g <sub>2</sub> , g <sub>3</sub> }	{3224, 2414}

CJ

CJ is the experience-shaping process by which consumers (c) use products and brands (p) to fulfill their goals [38]. Geometrically, each customer purchase behavior (c<sub>i</sub>, {p<sub>j</sub>}, {m<sub>j</sub>}, t) in Equation (1) is a scattering of points within the space  $\Xi(C,P,M,T)$ , but along the line (c<sub>i</sub>, t). In this space, a formal definition of CJ of the retail repurchase process can be formally derived.

**Definition 1**

Given a customer c<sub>i</sub>, his/her customer journey can be extracted from the transaction space as a triplet set:

$$\xi[c_i; \{(\{p_j\}, \{m_j\}, T)\} \subseteq 2^P \times 2^M \times T, \forall (c_i, \{p_j\}, \{m_j\}, t) \text{ with fixed } c_i, \text{ products (services) } \{p_j\}'s \subseteq P, \text{ expense } \{m_j\}'s \subseteq M, \text{ and time } t \in T$$

In Example 1,  $\xi_A[c_1] = \{(g_1, 1990, 2019/01/16), (g_1, 2682, 2019/06/02), (g_1, 3582, 2019/06/12), (g_1, 261, 2020/04/29), (g_4, 3944, 2020/05/08), (\{g_2, g_3\}, \{3224, 2414\}, 2020/07/06)\}$ . Here, each customer journey  $\xi[c_i]$  is defined as a series of customer purchase behaviors within the transaction space

**PM**

CJB

A CJ is a series of purchase behaviors of a customer. To measure CJ, some measurement duration should be defined first.

**Definition 2**

The CJB (CJ block)  $\zeta[c_i, T_k]$  is a series of measurement durations which transforms the time interval T into a series of time intervals [T<sub>k</sub>], in which T<sub>k</sub> is the beginning to some timestamp. Next, the set of customer journey blocks is defined as follows:

$$\Omega(C,P,M,T) = \{\zeta[c_i, T_k]\} = \{(\{p_j\}, \{m_j\}, t)\} \quad (2)$$

with products (services) {P<sub>j</sub>}’s  $\subseteq P$ , payment

$$\{m_j\}'s \subseteq M, \forall t \in T_k \subseteq T.$$

Such  $\zeta[c_i, T_k]$  is one kind of touchpoints (CJ building blocks), which collects the customer c<sub>i</sub>’s purchase experience within T<sub>k</sub>. CJ  $\xi[c_i] = \{\xi[c_i, T_k]\}$  is described as a series of touchpoints [26,28, 39].

For Example 1, as the time interval TA is segmented into [T<sub>1</sub>, T<sub>2</sub>, T<sub>3</sub>, T<sub>4</sub>] = [2019/01/01-2019/06/30, 2019/01/01-2019/12/31, 2019/01/01-2020/06/30, 2019/01/01-2020/12/31], then the aforementioned CJ  $\xi_A[c_1]$  can be segmented into

$$\xi_A[c_1, T_1] = \{(\{g_1, g_1, g_1\}, \{1990, 2682, 3582\})\}, \zeta_A[c_1, T_2] = \{(\{g_1, g_1, g_1\}, \{1990, 2682, 3582\})\}, \zeta_A[c_1, T_3] = \{(\{g_1, g_1, g_1, g_1, g_4\}, \{1990, 2682, 3582, 261, 3944\})\}, \text{ and } \zeta_A[c_1, T_4] = \{(\{g_1, g_1, g_1, g_1, g_4, g_2, g_3\}, \{1990, 2682, 3582, 261, 3944, 3224, 2414\})\}.$$

The aforementioned definitions of T<sub>k</sub>’s are all counted from the beginning, which ensures the product set is accumulated.

**LEMMA 1**

$$\forall T_i \subseteq T_j, \zeta[c_i, T_i] \subseteq \zeta[c_i, T_j].$$

**Proof**

Because  $T_i \subseteq T_j, T_{ji}$  exists such that  $T_i = T_i \cup T_{ji}$ , then,

$$\zeta[c_i, T_i] = \zeta[c_i, T_i \cup T_{ji}] = \zeta[c_i, T_i] \cup \zeta[c_i, T_{ji}] \supseteq \zeta[c_i, T_i].$$

PM

Irrespective of products (services) or time, a CJ presents an increment phenomenon, which indicates the quantity of its contents (products, services, frequencies or expenses) is increasing. For each CJ segment  $\zeta[c_i, T_{k0}]$ , an atomic PM  $\lambda(\{p_j\}, \{m_j\}, T_{k0})$  can be defined based on some quantity of the purchase behavior  $(\{p_j\}, \{m_j\}, t)$  in Equation (2) during customer c<sub>i</sub> purchasing within a basic time interval T<sub>k0</sub>, such as about:

- (1) the amounts / quantities of products (categories), services {p<sub>j</sub>},
- (2) the frequencies or time about t  $\in T_{k0}$ , or (3) the total sums of the expense {m<sub>j</sub>}.

Such measure satisfies two properties: (1)  $\lambda(\{p_j\}, \{m_j\}, T_{k0}) \geq 0$ , and

(2)  $\lambda(\{p_j\}, \{m_j\}, T_{k0}) = 0$ , if and only if  $\{(\{p_j\}, \{m_j\}, T_{k0})\}$  is an empty set, that is, no purchase within T<sub>k0</sub>. Therefore, a measure  $\mu: \Omega \rightarrow R + U\{0\}$  can be defined to summarize a description of products (services) or times in a CJ.

**Definition 3**

The PM of a CJ block  $\zeta[c_i, T_k]$  is a function mapping from a set of CJBs  $\Omega$  in Equation (2) to a non-negative real number, that is,

$$\mu: \Omega \rightarrow R^+ \cup \{0\}, \text{ and}$$

$$\mu(\zeta[c_i, T_k]) = \sum_{T_{k0} \subseteq T_k} \lambda(\{p_j\}, \{m_j\}, T_{k0}) \quad (3)$$

to aggregate all atomic PMs in this interval T<sub>k</sub>.

Applications of purchase measures (PMs)

For Example 1, several purchase measures (abbreviated as PMs)  $\mu(\cdot)$  in Equation (3) can be defined as follows. A PM  $\mu(\cdot)$  may be a real number, such as the purchase amount of products (services); for example,  $\mu_M(\zeta[c_1, \cdot]) = \{(1990+2682+3582=8254, T_1), (8254, T_2),$

$(8254+261+3944=12459, T_3), (12459+3224+2414=18079, T_4)\}$ . Similarly,

$\mu(\cdot)$  can be an integer to denote the purchase quantity or frequency.

The purchase quantity describes the purchased products (services), for example,  $\mu_Q(\zeta[c_1, \cdot]) = \{(3, T_1), (3, T_2), (5, T_3), (7, T_4)\}$ , and the frequency is a quantitative value of the visit times, for example,

$\mu_F(\zeta[c_i, \cdot]) = \{(3, T_1), (3, T_2), (5, T_3), (6, T_4)\}$ . By contrast,  $\mu(\cdot)$  in Equation (3) can be a logical value that indicates whether the customer purchases the product. For example,  $\mu_P(\zeta[c_i, T_3]) = [1 \ 0 \ 0 \ 1] = 1001$ . A PM can not only be a scalar but also be a vector (collection) of the purchased content, such as the aforementioned 10110, or precisely  $\mu_{PN}(\zeta[c_i, T_3]) = [4 \ 0 \ 0 \ 1] = 4001$ , in which the numbers indicate the quantities of the purchased products. Not restricted to the aforementioned products, numerous definitions of PMs have been proposed. The use of PM depends on its applications.

**PM space**

Definition of the PM space

Although PMs can be defined in various methods, they exhibit common properties and can be used to construct the purchase funnel.

**LEMMA 2**

A PM  $\mu(\cdot)$  is a (mathematical) measure.

**Proof**

A PM  $\mu(\cdot)$  satisfies the three properties of a mathematical measure: (1) non-negativity; (2) null empty set comes from the non-negativity of  $\lambda(\cdot)$ , and the vanishing of  $\lambda(\cdot)$  represents no purchase behavior; (3) "countable additivity" of a mathematical measure can be derived from the following equation:

$$\begin{aligned} \mu(\zeta[c_i, T_{k1} \cup T_{k2}]) &= \sum_{T_{k0} \subseteq T_{k1} \cup T_{k2}} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &= \sum_{T_{k0} \subseteq (T_{k1} \setminus T_{k2}) \cup (T_{k2} \setminus T_{k1}) \cup (T_{k1} \cap T_{k2})} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &= \sum_{T_{k0} \subseteq (T_{k1} \setminus T_{k2})} \lambda(\{p_j\}, \{m_i\}, T_{k0}) + \sum_{T_{k0} \subseteq (T_{k2} \setminus T_{k1})} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &\quad + \sum_{T_{k0} \subseteq (T_{k1} \cap T_{k2})} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &= \mu(\zeta[c_i, T_{k1} \setminus T_{k2}]) + \mu(\zeta[c_i, T_{k2} \setminus T_{k1}]) + \mu(\zeta[c_i, T_{k1} \cap T_{k2}]). \end{aligned}$$

Mathematically,  $\Xi$  forms a transaction space and its customer journey  $\{\zeta[c_i, T_k] = \{(p_j, \{m_i\}, t)\}$  forms a measurable space  $(\Xi, \zeta)$ . Accordingly, PM  $\mu(\cdot)$  is a (mathematical) measure on  $(\Xi, \zeta)$ . The triple  $(\Xi, \zeta, \mu)$  forms a measure space, called *purchase measure space* (PM space).

Properties of PM space

Because PM space  $(\Xi, \zeta, \mu)$  is a standard measure space, a set of basic properties can be derived accordingly.

**Proposition 1**

Defined in the PM space  $(\Xi, \zeta, \mu)$ , the PM  $\mu(\cdot)$  exhibits the following basic properties of measure space:

- (1) Monotonicity:  $\mu(\zeta[c_i, T_{k1}]) \leq \mu(\zeta[c_i, T_{k2}])$  for  $T_{k1} \subseteq T_{k2}$ .
- (2) Subadditivity:  $\mu(\cup_i \zeta[c_i, T]) \leq \sum_i (\mu(\zeta[c_i, T]))$ .

**Proof**

(1) By Lemma 1, the inclusion of  $T_{k1} \subseteq T_{k2}$  reveals that  $(p_j, m_i)$  exists such that  $\zeta[c_i, T_{k2}] = \zeta[c_i, T_{k1}] \cup \zeta[c_i, T_{j,k1}]$ , then we have the

following expression:

$$\begin{aligned} \mu(\zeta[c_i, T_{k2}]) &= \sum_{T_{k0} \subseteq T_{k1} \cup T_{j,k1}} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &= \sum_{T_{k0} \subseteq T_{k1}} \lambda(\{p_j\}, \{m_i\}, T_{k0}) + \sum_{T_{k0} \subseteq T_{j,k1}} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &\geq \sum_{T_{k0} \subseteq T_{k1}} \lambda(\{p_j\}, \{m_i\}, T_{k0}) \\ &= \mu(\zeta[c_i, T_{k1}]) \end{aligned}$$

by Lemma 2.

(2) By countable additivity,

$$\begin{aligned} \mu(\zeta[c_i, T_{k1} \cup T_{k2}]) &= \mu(\zeta[c_i, T_{k1}]) + \mu(\zeta[c_i, T_{k2}]) - \mu(\zeta[c_i, T_{k1} \cap T_{k2}]) \\ &\leq \sum_{j=1,2} \mu(\zeta[c_i, T_{kj}]). \end{aligned}$$

Next, the subadditivity property can be obtained by induction.

For a standard measure space, many other properties, such as continuity from below/above, completeness, additivity, and so on, are not used in our derivation and neglected here.

Increment of PMs

By the monotonicity of the PM  $\mu(\cdot)$  along a specific customer journey  $\zeta[c_i, T_k]$ , the *increment of PMs* can be defined as follows:

$$\delta(\mu(\zeta[c_i, T_k]) = \mu(\zeta[c_i, T_k]) - \mu(\zeta[c_i, T_{k-1}]), \tag{4}$$

which denotes the incremental quantity of the PM in touchpoint  $\zeta[c_i, T_{k-1}]$  to the next touchpoint  $\zeta[c_i, T_k]$ . To complete the definition, the initial value is set as  $\delta(\mu(\zeta[c_i, T_1]) = \mu(\zeta[c_i, T_1])$ .

From Equation (4), the aforementioned PMs  $\mu_M(\cdot)$  can provide  $\delta(\mu_M(\zeta[c_i, \cdot])) = \{(8254 - 0 = 8254, T_1), (8254 - 8254 = 0, T_2), (12459 - 8254 = 4205, T_3), \text{ and } (19079 - 12459 = 5638, T_4)\}$ . Moreover,

$$\begin{aligned} \delta(\mu_P(\zeta[c_i, \cdot])) &= \{(3, T_1), (0, T_2), (2, T_3), (2, T_4)\}, \delta(\mu_F(\zeta[c_i, \cdot])) = \{(3, T_1), (0, T_2), (2, T_3), (1, T_4)\}, \\ \delta(\mu_P(\zeta[c_i, \cdot])) &= \{(1000, T_1), (0000, T_2), (1001, T_3), (0110, T_4)\}, \end{aligned}$$

$$\delta(\mu_{PN}(\zeta[c_i, \cdot])) = \{(3000, T_1), (0000, T_2), (1001, T_3), (0110, T_4)\}.$$

In Section 3, purchase measures, first, are used to construct the PMG. Next, in Section 4, the increment of purchase measures is formally used to investigate the quantitative properties of the PM in CJ.

**PMG AND ITS KNOWLEDGE ANALYSIS**

**PM quotient space**

PM equivalence relation

If values of PM  $\mu(\cdot)$  are finite, define a *PM equivalence relation* is defined as  $R_\mu(x_1, x_2) = \{\mu(x_1) = \mu(x_2) : x_1, x_2 \in \Xi\}$ . For the purchase product measure  $\mu_P(\cdot)$ , a more all-zero-products *pseudo-source* (0000 in Example 1) is added to treat all customers as new comers. If not, more elaborate designed can be detailed. Thus,  $\mu_P(\cdot)$  provides the PM space with  $n(C) * (n(T)+1)$  values, where  $n(C)$  and  $n(T)$  are the sizes of customer set  $C$  and time interval set  $T$ , respectively.

For  $\mu_P(\cdot)$  in Example 1, although  $5*(4+1)=25$  measure values are possible, and the customer  $c_1$  has only five elements:

$$\mu_P(\zeta_A[c_1, T_0-T_4]) = \{0000, 1000, 1000, 1001, 1111\}.$$

For the expense measure  $\mu_M(\cdot)$ , quantity measure  $\mu_Q(\cdot)$ , frequency measure  $\mu_F(\cdot)$ , and the product quantity measure  $\mu_{PN}(\cdot)$  are used. Infinite measure values are possible and they should be quantified.

PM equivalence class, PM quotient space

With the finite or quantified measure values and their resultant equivalence relation, the equivalence class of PM space and the corresponding quotient space can be defined.

**Definition 4**

Given the PM space  $(\Xi, \zeta, \mu)$  with finite-valued measure  $\mu(\cdot)$ , a PM equivalence class of  $(\Xi, \zeta, \mu)$  can be defined as follows:

$$[q_k] = \{x \in \Xi, \mu(x) = q_k\} = \Xi / R_\mu = Q_\mu, \tag{5}$$

which is the PM quotient space (set)  $Q_\mu$  of  $\Xi$  and each  $[q_k]$  is a partition of space .

Following the PM  $\mu_P(\cdot)$  of  $\zeta_A[\cdot, \cdot]$  in Example 1, the corresponding quotient space in Equation (5) can be  $Q_{\mu_P} = \{[0000], [1000], [1001], [1011], [1100], [1110], [1111]\}$ , in which  $[1000] = \{[c_1, T_2], [c_1, T_3], [c_2, T_3], [c_2, T_4]\}$  with size 4.

Size of the PM quotient space

Similarly, the count of elements in partition  $Q_\mu$  of Equation (5) is denoted as  $n(Q_\mu)$  and is called the size of the PM quotient space, which satisfies the condition:  $\sum_k (n(Q_\mu = [q_k]) = n(\zeta))$ . For example,  $n(Q_{\mu_P}[\zeta_A]) = \{11, 4, 1, 3, 3, 1, 2\}$ , whose sum just match the number of elements in  $\zeta_A$ ,  $n(\zeta_A) = n(C) * (n(T)+1) = 5 * (4+1) = 25$ .

**PMG**

PM relation

Next, the relation between the elements of PM equivalence classes, which comes from CJ, are evaluated.

**Definition 5**

Given the PM space and its PM Equivalence Class (EC)

$$Q_\mu = [q_k] = \{x \in \Xi(C, P, M, T), \mu(x = \zeta[c_i, T_k]) = q_k\}, \text{ a PM relation is}$$

$$E(\Xi, \zeta, \mu) = \{(q_i, q_j)\} = \{(\mu(\zeta[c_i, T_k]), \mu(\zeta[c_i, T_{k+1}])) : c_i \in C, T_k \subseteq T\} \subseteq Q_\mu \times Q_\mu. \tag{6}$$

In graph theory, quasi-ordinal is defined as a relation that satisfies reflexivity and transitivity. With this definition, the following theorem can be obtained easily.

**Theorem 1**

Given the PM space, the PM relation  $E(\Xi, \zeta, \mu)$  is a quasiordinal relation.

**Proof**

From Lemmas 1,2 and Proposition 1,  $T_k \subseteq T_{k+1}$  makes  $\zeta[c_i, T_k] \subseteq \zeta[c_i, T_{k+1}]$ , and  $\mu(\zeta[c_i, T_k]) \leq \mu(\zeta[c_i, T_{k+1}])$ , then the PM relations of  $E(\Xi, \zeta, \mu) = \{(q_i, q_j)\}$  is quasi-ordinal.

PMG

The aforementioned PM EC collects the touchpoints of CJs with similar PM  $\mu(\cdot)$ . Therefore, purchase funnel can be defined as a partial ordering graph.

**Definition 6**

Given the PM space  $(\Xi, \zeta, \mu)$  and its PM EC  $Q_\mu = [q_k]$  satisfying Equation (5), a PMG can be defined as  $\Psi(Q_\mu, E)$ , where the vertex set is the PM EC in Equation (5),  $V_{set}(\Psi) = Q_\mu$ , and the edge set is  $E_{set}(\Psi) = E(\Xi, \zeta, \mu)$  in Equation (6).

Combining previous derivation, such PMG can be obtained through the following algorithm.

**Algorithm 1. (PMG Algorithm)**

Given a PM space, the PMG  $\Psi(Q_\mu, E)$  can be obtained by the following steps:

- (1) Find all the customer journeys in Equation (2)  $\zeta[c_i, T_k] = (\{p_j\}, \{m_j\}, t), \forall c_i \in C \text{ and } \forall T_k \subseteq T$ .
- (2) Add pseudo-source to each customer and calculate all the PMs  $\mu(\zeta[c_i, T_k])$  in Equation (3).
- (3) Divide the measure space  $\Xi$  into several ECs  $Q_\mu$ 's in Equation (5), which also forms the vertex set  $V_{set}$ .
- (4) Find the edge set  $E_{set}$  in Equation (6) from the consecutive PMs  $(\mu(\zeta[c_i, T_k]), \mu(\zeta[c_i, T_{k+1}]))$ 's in all the CJs.
- (5) Then form the PMG  $\Psi(Q_\mu, E) = \text{graph}(V_{set}, E_{set})$ .

Because PM measure  $\mu(\cdot)$  satisfies the property of monotonical increasing, as presented in Proposition 1, the resultant PMG exhibits the partial ordering property.

PMG for example 1

For the construction of PMG in Example 1, the vertex set in Equation (5) is adopted as

$V_A = Q_{\mu_{PA}} = \{[0000], [1000], [1001], [1011], [1100], [1110], [1111]\}$  and edge set E in Equation (6) can be obtained from the collection of CJs of all customers, such as  $0000 \rightarrow 1000 \rightarrow 1001 \rightarrow 1111$  from customer  $c_1$ . The resultant PMG is displayed in Figure 2.

In Figure 2, the CJ block in each vertex (EC) is shortened in a notation sum  $c_i/T_k$ , which indicates the aggregation of  $[c_i, T_k]$ . For the aforementioned example of  $[1000] = \{[c_1, T_2], [c_1, T_3], [c_2, T_3], [c_2, T_4]\}$  is denoted as  $1000 = 1/2 + 1/3 + 2/3 + 2/4$  in Figure 2.

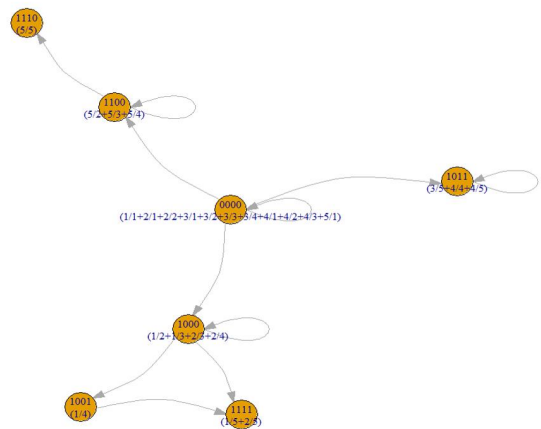


Figure 2) PMG of Example .

Pseudo-source design

With the pseudo-source design in Section 3.1, all the nodes in the

PMG  $\Psi(Q_\mu, E)$  originate from the pseudo-source. As displayed in Figure 2, pseudo-source (0000) can integrate all the three components of  $\Psi(Q_\mu, E)$  into one connected graph and causes PMG to exhibit well-behaved properties.

**PM knowledge structure**

With the partial ordering of PMGs, knowledge space theory (KST) can be used to study the CJ [40].

The aforementioned set of CJs  $\zeta = \{\zeta_i = \zeta[c_i]\} \subseteq 2^{Q_\mu}$  is the domain to be processed, such as  $\zeta[c_1] = \{0000, 1000, 1001, 1111\}$ . Therefore, the basic KST and knowledge structure can be introduced with the PM.

**Definition 7**

Given a PMG  $\Psi_f(Q_\mu, E_f)$ , a *purchase measure knowledge structure*, is defined as  $(\zeta, K)$ , where  $\zeta$  is the domain of CJs and the set of knowledge states  $K \subseteq 2^\zeta$  and  $K \neq \emptyset$  (empty set).

In Example 1,  $\zeta_A = \{\{0000,1000,1111\}, \{0000, 1011\}, \{0000,1100,1110\}, \{0000,1000,1001,1111\}\}$ , each of which corresponds to the CJs of  $\{c_2\}$ ,  $\{c_3, c_4\}$ ,  $\{c_3\}$  and  $\{c_1\}$ . Notably,  $\zeta_A \subseteq 2^{Q_{\mu A}}$ . Moreover, a subset of  $\zeta_A$  such as  $\{0000\} \rightarrow \{0000,1000\} \rightarrow \{0000,1000,1111\}$  is an element of the PM knowledge structure  $K_A$ , that is,  $K_A \subseteq 2^{\zeta_A}$ .

Quasi-ordinal knowledge structure

When a knowledge structure  $K$  is closed under intersection, such knowledge structure  $(\zeta^+, K)$  becomes a quasi-ordinal relation. To make  $(\zeta, K)$  quasi-ordinal knowledge structure, those knowledge states  $\zeta_i^+$ 's  $\in \zeta^+$ , where

$$\zeta_i^+ = \{\zeta_i \cap \zeta_j : \zeta_i, \zeta_j \in \zeta\} \tag{7}$$

encircles not only  $\zeta$ , but their partial paths, that is, the subsets, of  $2^{Q_\mu}$ . Such knowledge structure extends the domain set from  $\zeta$  to  $\zeta^+$ .

For example, in addition to  $\zeta_A$ ,  $\zeta_A^+$  still contains  $\{0000, 1000\}$ ,  $\{0000, 1000, 1111\}$  and the single pseudo-source  $\{0000\}$ . Moreover, such intersection closeness condition of can be described formally.

**LEMMA 3**

For one quasi-ordinal PM knowledge structure in Equation (7),

$$\zeta_i \cap \zeta_j \in \zeta^+, \forall \zeta_i, \zeta_j \in \zeta^+.$$

**Proof**

By definition.

**PM knowledge space**

In mathematics, the construction of a space represents the following: (1) finding the basic elements of this space is possible. Such elements are typically named bases. (2) All the elements in this space can be spanned by the combinations of these "bases." The spanning logic is always illustrated by a set of "coordinates".

**Definition 8**

A PM knowledge structure  $(\zeta^+, K)$  is called a PM knowledge space

$$(\zeta^+, K^*), \text{ where } K^* \text{ is closed under union, that is, } K^* = \{\zeta_i \cup \zeta_j \in \zeta^+, \forall \zeta_i, \zeta_j \in \zeta^+\}. \tag{8}$$

Such union is a coordinate, which is explained subsequently. The following property occurs naturally.

**Proposition 2**

Given a PMG  $\Psi_f(Q_\mu, E_f)$ , the purchase measure knowledge space

$$(\zeta^+, K^*), \text{ satisfies } (\zeta^+, K^*) \supseteq (\zeta, K), \text{ which reveals that } \zeta^+ \supseteq \zeta \text{ and } K^* \supseteq K.$$

**Proof**

By definition.

With union operations of knowledge states in Example 1, its knowledge space  $(\zeta_A^+, K_A^*)$  can be obtained as follows:

**PM knowledge base**

According to research, a knowledge space  $\Gamma(\zeta^+, K^*)$  from Equations (7) and (8) admit at most one knowledge base, called PM knowledge base  $B_K$ , which is a minimal family of knowledge states spanning the knowledge space [41]. Such a knowledge base can be found by an "algorithm for constructing knowledge base" and is implemented in a ready-made software [42].

$$B_K = \{\beta_k\} = \text{kbases}((\zeta^+, K^*)). \tag{9}$$

Based on previous research, the PM knowledge base of the PM

$$\text{knowledge space in Example 1 can be } B_{K_A} = \text{kbases}((\zeta_A^+, K_A^*)) = \{\{0000\}, \{0000, 1000\}, \{0000, 1011\}, \{0000, 1100, 1110\}, \{0000, 1000, 1001, 1111\}\}.$$

CJs and future prediction

From the perspective of CJ, knowledge base may provide considerable insights. First, the pseudo-source (e.g., 0000) is the staying point of the new customer. Other knowledge bases function as the purchase-starting points of CJ tracks.

In Figure 3 of Example 1, 0000-1000, 0000-1011, and 0000-1100 are the starting points of CJs  $\{c_1, c_2\}$ ,  $\{c_3, c_4\}$ ,  $\{c_3\}$ , respectively. Notably, the length-2 purchase path 0000-1000-1001-1111 is an extended purchase of 0000-1000 and plays the role of a knowledge base, which is investigated in the following section.

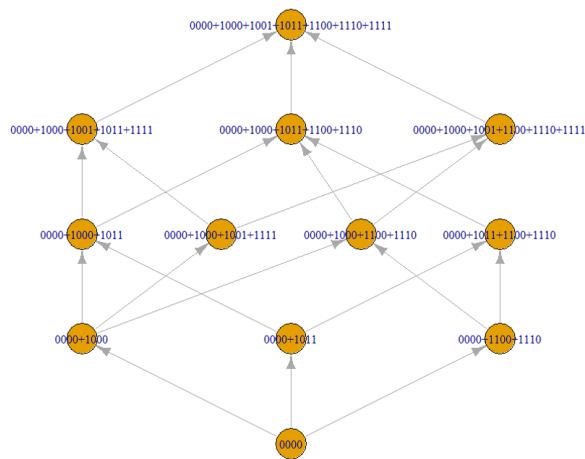


Figure 3) PM knowledge space of example 1

CJG IN THE PM KNOWLEDGE SPACE

PM coordinate system and CJG

PM coordinate system

Before spanning the knowledge space by knowledge bases, investigating the knowledge bases is necessary. From the perspective of the coordinate vector space, a basis is a set of independent vectors  $\{\beta_k\}$ , such as Equation (9), called PM axes, to completely describe (span) the vector space and the PM origin  $\beta_0$  is the intersection of these bases.

In the PM knowledge base of Example 1,  $\beta_{kA1} = 0000 - 1000$ ,  $\beta_{kA2} = 0000 - 1011$  and  $\beta_{kA3} = 0000 - 1100$  are three independent PM vectors (axes), and  $\beta_{kA0} = 0000$  is the origin. Moreover, the PM knowledge base has the PM path of the path length more than one, such as  $\beta_{kA1*} = 0000 - 1000 - 1001 - 1111$ , which should be in the direction of  $\beta_{kA1} = 0000 - 1000$ .

Therefore, the PM coordinate system can be organized as follows.

Algorithm 2. (Algorithm for PM coordinate system)

The coordinate system  $A_k = \{a_k\}$  with respect to (w. r. t.) PM knowledge base  $B_k = \{\beta_k\}$  in Equation (9) can be coded as follows:

1. Coordinate w.r.t. pseudo-source  $\beta_0$ : coded as the origin;
2. Coordinate w.r.t. PM axes  $\beta_k$ : coordinate as a unit dimensional vector  $a_k$ ,
3. Extended coordinate w.r.t. purchase path extended axes  $\beta_k$ : of path length more than one: coded as the dimensional vector of the path length along axes.

Using this algorithm, the aforementioned PM knowledge base  $B_{kA} = \{0000, \{0000, 1000\}, \{0000, 1011\}, \{0000, 1100, 1110\}, \{0000, 1000, 1001, 1111\}\}$  can be coded as  $A_{kA} = \{[0,0,0], [1,0,0], [0,1,0], [0,0,1], [2,0,0]\}$ .

PM knowledge space spanning

With knowledge base  $B = \{\beta_k\}$  in Equation (9), it is sufficient to reconstruct the full knowledge space as follows:

$$K^* = \{\cup \beta_k : \beta_k \in B_k\}. \tag{10}$$

With PM knowledge base  $B = \{\beta_k\}$  and its corresponding PM coordinate systems  $A = \{a_k\}$ , the PM knowledge space  $(\zeta^+, K^*)$  in Equation (7) and (8) can be easily constructed by  $\zeta^+$  as the using domain of  $K^*$  [41].

From the perspective of evolution theory, when customers exchange their purchased information with each other, it causes purchase behavior interaction, which results in mating or mutation in evolution theory. In terms of the knowledge space, a union operation is formed between purchase behaviors, which can be used to predict possible future knowledge states.

CJG

By the partial ordering of the subset property of  $K^*$ , the PM knowledge space in Figure 3 is a graph of the Hasse diagram, called CJG.

Definition 9

Given a PMG  $\Psi(Q, \mu, E)$  and its PM knowledge space  $(\zeta^+, K^*)$  with CJ block in Equation (7) and PM knowledge space  $K^*$  in Equation (8), then a CJG of  $(\zeta^+, K^*)$  is defined as  $\Gamma(K^*, \Lambda)$ , where the vertex set is the PM knowledge space, that is,  $Vset(\Gamma) = (\zeta^+, K^*)$ , and the edge set is

$$Eset(\Gamma) = \{(\zeta_{k1}, \zeta_{k2}) : \mu(\zeta_{k1}) < \mu(\zeta_{k2})\} \equiv \Lambda. \tag{11}$$

The resultant CJG  $\Gamma_A(K_A^*, E_A)$  in Example 1 is displayed in Figure 4, in which each circle keeps a coordinate (000 to 211) and its corresponding customer journey (0000 to 0000, 1000, 1001, 1011, 1100, 1110, 1111).

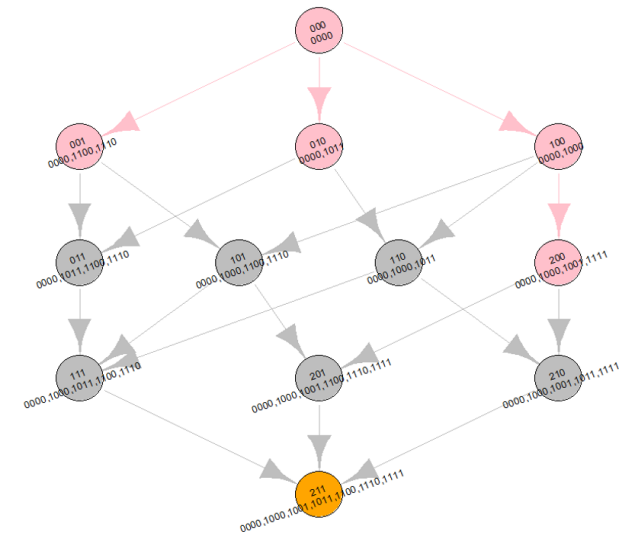


Figure 4) CJG of Example 1

By knowledge space analysis, these nodes  $K_A^*$  in Figure 4 not only represent CJs but also cover the PM knowledge states, which have coordinates in the form of  $a_k = (x_k, y_k, z_k)$ , where  $x_k = 0 - 2$  and  $y_k, z_k = 0 / 1$ . Thus, the overall PM knowledge space is a three-dimensional (3D) cube with  $3 \times 2 \times 2 = 12$  elements. The x-axis has two unit vectors,  $0000 - 1000$  and  $0000 - 1000 - 1001 - 1111$ , coordinated as (1,0,0) and (2,0,0). The unit vectors of y-axis and z-axis are  $0000 - 1011$  (with coordinate (0,1,0)) and  $0000 - 1100 - 1110$  (with coordinate (0,0,1)), respectively. These knowledge states in the



knowledge base or the knowledge structure are represented as pink or orange circles, respectively.

Furthermore, the rest knowledge states shown in gray circles, for example, 011 (0000,1011,1100,1110) and 201 (0000,1000,1001,1100,1110,1111), are not in any dimensional axis and exhibit the coordinates of more than one digit. Such knowledge states denote mixed CJs and represent the future possible consumption behaviors of these customers.

**Volumes and flows in the PM knowledge space**

CJ code

Each node of CJG  $\Gamma(K^*, \Lambda)$  of Equation (11) is a knowledge state  $\zeta_i$  (or  $\mu(\zeta_i)$ ) precisely of the knowledge space, which denotes just a customer journey of some customers, such as  $\zeta_A [c_1] = 0000 - 1000 - 1001 - 1111$ . In this example,  $\mu(\zeta)$  are the nodes in the PMG  $\Psi_f(Q_\mu, E_f)$ , that is,

$$Vset(\Psi_A(Q_{\mu A}, E_{fA})) = [0000, 1000, 1001, 1011, 1100, 1110, 1111],$$

representing purchase product patterns (or more precisely product category patterns) at one purchase duration.

By contrast, one customer journey  $\zeta_i$  in a knowledge state, such as  $\zeta_A [c_1]$ , is one purchase pattern, which describes the purchase sequence (0000, 1000, 1001, 1111) along this CJ. Such knowledge state can then be coded as a 0/1-vector corresponding to one purchase product pattern in  $Vset(\Psi(Q_\mu, E))$ .

**Definition 10**

Given a PMG  $\Psi_f(Q_\mu, E_f)$  and its corresponding knowledge space  $(\zeta^+, K^*)$ , then for each customer journey (knowledge state)  $\zeta_i \in K^*$ , its corresponding CJ code (along PMG) is defined as follows:

$$\eta_i = \eta(\zeta_i) = [\eta_{ik}(\zeta_i), k = 1, \dots, n(Q_\mu)] \tag{12}$$

where  $\eta_{ik} = 1$  if  $Q_{\mu k} \in \zeta_i$  and  $\eta_{ik} = 0$  if elsewhere.

For example,  $\eta_{A1} = \eta(\zeta_A [c_1]) = [1, 1, 1, 0, 0, 0, 1]$  or 1110001 and all the nodes in CJG  $\Gamma_A(K_A^*, E_A)$  can be coded as 1000000 (0000), 1100000 (0000,1000), ..., 1111111 (0000,1000,1001, 1011,1100,1110,1111), which are listed in Table 2, called the codebook of CJ knowledge space. In the first node, the first element in the top node 1000000 indicates 0000, which is pseudo-source and included in all  $\zeta$ 's nodes.

**TABLE 2**  
**Codebook of the CJ knowledge space in Example 1**

	0	1000	1001	1011	1100	1110	1111
0	1	0	0	0	0	0	0
0000,1000	1	1	0	0	0	0	0
0000,1011	1	0	0	1	0	0	0
0000,1000,1011	1	1	0	1	0	0	0
0000,1100,1110	1	0	0	0	1	1	0
0000,1000,1001,1111	1	1	1	0	0	0	1

0000,1000,1100,1110	1	1	0	0	1	1	0
0000,1011,1100,1110	1	0	0	1	1	1	0
0000,1001,1000,1011,1111	1	1	1	1	0	0	1
0000,1000,1011,1100,1110	1	1	0	1	1	1	0
0000,1000,1001,1100,1110,1111	1	1	1	0	1	1	1
0000,1000,1001,1011,1100,1110,1111	1	1	1	1	1	1	1

Volume of the knowledge state

To investigate the aforementioned CJG with the node set in Equation (10) and edge set in Equation (11), the volume of a knowledge state  $\zeta_k \in K^*$  of  $(\zeta^+, K^*)$  is defined as the number of customers in this state  $\zeta_k$ , symbolized as follows:

$$volume(\zeta_k) = \{n(c_i) : \exists T_k, \zeta_k = \zeta[c_i, T_k]\}. \tag{13}$$

For  $(\zeta_A^+, K_A^*)$  in Example 1,  $volume([1000000, 1100000, 1001000, 1101000, 1000110, 1110001]) = [5, 2, 2, 0, 1, 2]$  and the volumes of the other nodes are all zeros. The top node pseudo-source  $\beta_0$  possesses the volume  $n(C)$ , which indicates that all CJs start from the very node.

Flow between knowledge states

When a customer  $c_i$  makes a purchase  $\mu(\zeta[\cdot])$  in the buying process, the CJ (knowledge state  $\mu(\zeta[c_i, T_k])$ ) will switch to another CJ (knowledge state  $\mu(\zeta[c_i, T_{k+1}])$ ). For customer  $c_1$ , a more purchase of  $g_1$  in  $T_1$  renders  $\zeta[c_1, \cdot]$  from CJ 0000 (knowledge state 1000000) to CJ 0000,1000 (state 1100000). Such knowledge state switching also occurs in  $\zeta[c_2, T_2]$ , which causes an edge flow of 2. Formally, we have the following expression:

$$flow(\zeta_1, \zeta_2) = \{n(c_i) : \exists T_{k-1}, T_k, \zeta_1 = \zeta[c_i, T_{k-1}] \text{ and } \zeta_2 = \zeta[c_i, T_k]\} \tag{14}$$

for  $(\zeta_1, \zeta_2) \in \Lambda$  in Equation (11).

**Quantitative analysis in CJG**

Gradation of CJ

For a CJ  $\zeta[c_i, \cdot]$ , gradation is defined as the difference in the touchpoint  $\zeta[c_i, T_k]$  from its predecessor  $\zeta[c_i, T_{k-1}]$  by one item [43]. Equation (4), the gradation represents the vanishing or unit quantity of increment PM as follows:

$$n(\delta(\mu(\zeta[c_i, \cdot]))) = 0 \text{ or } 1, \tag{15}$$

Which indicates  $\zeta[c_i, T_k]$  is the same as the predecessor  $\zeta[c_i, T_{k-1}]$  (0) or just purchase one-more single product/category (1).

The gradation of CJ indicates that a purchase CJ is a stepwise customer journey, which indicates a single purchase in one touchpoint (CJB), not multiple purchases. The descriptions about PM increments of Example 1 reveal that gradation is not common in customer journey; however, this condition can hold under certain conditions and is highly characterized on the purchase measure.

CJG and its quantities

The volumes and switching flow of the knowledge states were merged into the CJG, whose nodes are encoded as customer journey codes, as displayed in Figure 5. In CJG graph, the paths from the pseudo-source to the terminal nodes are the purchase journeys of customers. Each node in the CJG is a knowledge state and denotes a purchase journey of the customer, and the node label is in the form of CJ coding (e.g., 110000) followed by its volume in a parenthesis (e.g., (2)), and the purchased product categories are listed in the following diagram (e.g., 0000,1000).

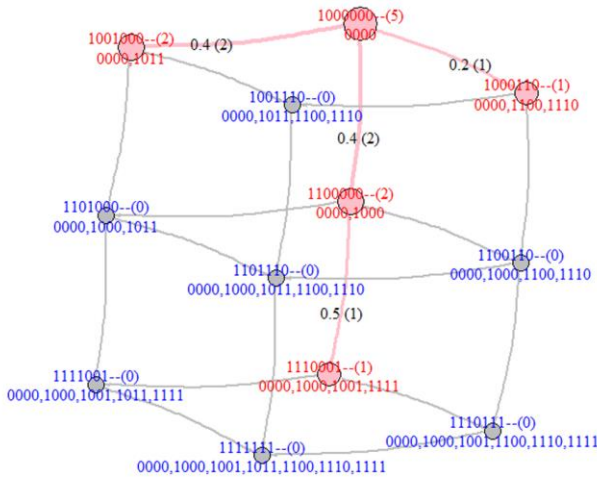


Figure 5) Quantified CJG of example 1

Quantitative properties of CJG

A CJG includes two knowledge nodes: (1) the nodes in knowledge bases: displayed in red circles in Figure 5, and (2) the nodes in the knowledge space but not in the knowledge base: displayed as gray circles in Figure 5. The edges and paths between two CJ nodes indicate the switching between these two nodes.

Such knowledge state switching exhibits filtering properties in the sense of product purchase and customer volume.

**LEMMA 4**

Given a customer journey graph  $\Gamma(K^*, \Lambda)$ , the nodes (customer journeys) exhibit two filtering properties: (1) the increasing in PMs :  $\mu(\zeta_{j1}) \leq \mu(\zeta_{j2}), \forall (\zeta_{j1}, \zeta_{j2}) \in \Lambda$ ; and moreover, for all  $Eset(\Gamma) = \Lambda$  of gradation customer journey graph, (2) the decrease in customer volumes :  $volume(\zeta_{j1}) \geq volume(\zeta_{j2}), \forall (\zeta_{j1}, \zeta_{j2}) \in \Lambda$ .

**Proof**

(1) By the definition of

$$\Lambda = Eset(\Gamma) = \{(\zeta_{j1}, \zeta_{j2})\} \equiv \Lambda.(2)\forall (\zeta_{j1}, \zeta_{j2}) \in \Lambda, \exists c_i \text{ and}$$

$T_k, \zeta_{j1} = \zeta[c_i, T_{k-1}]$  and  $\zeta_{j2} = \zeta[c_i, T_k]$  to make an edge of Eset of  $\Gamma$ .

By the gradation property of CJ  $\zeta[c_i, T_k]$ ,

$$volume(\zeta_{j1} = \zeta[c_i, T_{k-1}]) - volume(\zeta_{j2} = \zeta[c_i, T_k]) =$$

$$\{n(c_i) : n(\delta(\mu(\zeta[c_i, T_k]))) = 0 \text{ or } 1\} = \{n(c_i) : n(\delta(\mu(\zeta[c_i, T_k]))) = 1\} \geq 0.$$

Although Lemma 4 can be proven only for gradation CJs, the decreasing property of customer volumes holds in almost all the

conditions.

Conversion rate and CJG algorithm

Based on the node volume in Equation (13) and edge flow in Equation (14), a meaning term, conversion rate, can be defined for each edge  $(\zeta_{k1}, \zeta_{k2}) \in \Lambda$  as follows:

$$conversion(\zeta_{k1}, \zeta_{k2}) = volume(\zeta_{k2}) / volume(\zeta_{k1}). \tag{16}$$

With the node volume and edge conversion rate, the aforementioned CJG such as Figure 5 can be developed in a systematic way.

**Algorithm 3. (CJG algorithm)**

Given the PM space  $(\Xi, \zeta, \mu)$ , the customer journey graph (CJG)

$\Gamma(K^*, \Lambda)$  can be obtained by the following steps:

- (1) Construct the PMG  $\Psi(Q_\mu, E)$  using Algorithm 1.
- (2) Build up the quasi-ordinal knowledge structure of  $(\zeta^+, K)$  from Equation (7) and Equation (8) with the intersection closure.
- (3) Furthermore, find the knowledge space  $(\zeta^+, K^*)$  with union closure.
- (4) Find the corresponding knowledge base  $B_K = \{\beta_k\} = kbase((\zeta^+, K^*))$  in Equation (9).
- (5) Let the vertex set  $K^* = \{\zeta_i\}$  in Equation (10) be the knowledge states of the knowledge space  $(\zeta^+, K^*)$ , then  $VL = \{volume(\zeta_k)\}$  from Equation (13).
- (6) For  $\zeta_{k1}, \zeta_{k2} \in \Lambda = \{(\zeta_{k1}, \zeta_{k2}) : \mu(\zeta_{k1}) < \mu(\zeta_{k2})\}$  in Equation (11), calculate flow  $FL = \{flow(\zeta_1, \zeta_2)\}$  from Equation (14), and conversion rate  $CR = \{conversion(\zeta_{k1}, \zeta_{k2})\}$  from Equation (16).
- (7) Then the CJG is found as CJG  $\Gamma(K^*, \Lambda, VL, FL, CR)$ .

CJG filtering

For Example1, the CJ code of pseudo-source in Figure 5 is 1000000. The path 1000000  $\rightarrow$  1000110 represents customer journey  $\zeta\{c_5, \cdot\}$  and the path 1000000  $\rightarrow$  1001000 includes  $\zeta\{c_3, \cdot\}$  and  $\zeta\{c_4, \cdot\}$ . Then these two paths own flows 1 and 2, respectively, and exhibit conversion rates  $1/5=0.2$  and  $2/5=0.4$ , respectively. However, the path 1000000  $\rightarrow$  1100000 encircles  $\zeta\{c_1, \cdot\}$  and  $\zeta\{c_2, \cdot\}$ , and the extended path 1100000  $\rightarrow$  1110001 has only one customer  $c_1$ . Then, the corresponding flows are 2 and 1, respectively, and the conversion rates are  $2/5=0.4$  and  $1/2=0.5$ , respectively.

This example indicates a direct corollary of Theorem 2. This CJG can also be called a purchase funnel, which occurs from the filtering properties of the CJG.

**Theorem 2. (CJG Filtering of the purchase funnel)**

Given a customer journey graph  $\Gamma(K^*, \Lambda, V, F, R)$  constructed using Algorithm 3, the node volumes  $V$  in Equation (13) and edge

conversion rates R in Equation (16) exhibit the following relationships: (1)  $\Sigma_{\zeta_{j2}}$  conversion  $(\zeta_{j1}, \zeta_{j2}) \leq 1$  for any node  $\zeta_{j1} \in K^*$ .

(2)  $\text{volume}(\zeta_j) = \text{volume}(\zeta_i) * \Pi_{\zeta_{k1}, \zeta_{k2}}$  conversion  $(\zeta_{k1}, \zeta_{k2})$  with the first  $\zeta_{k1} = \zeta_i$  and the last  $\zeta_{k2} = \zeta_j$ .

**Proof:**

- (1)  $\Sigma_{\zeta_{j2}}$  conversion  $(\zeta_{j1}, \zeta_{j2}) = \Sigma_{(\zeta_{j1}, \zeta_{j2}) \in \text{Eset}(\Gamma)}$   
 $\text{volume}(\zeta_{j2}) / \text{volume}(\zeta_{j1}) = \Sigma_{(\zeta_{j1}, \zeta_{j2}) \in \text{Eset}(\Gamma)} \text{volume}(\zeta_{j2}) / \text{volume}(\zeta_{j1}) =$   
 $= \Sigma_{(\zeta_{j1}, \zeta_{j2}) \in \text{Eset}(\Gamma)} \{n(c_i) : \mu(\zeta[c_i, \dots]) = \mu(\zeta_{j2})\} / \{n(c_i) : \mu(\zeta[c_i, \dots]) = \mu(\zeta_{j1})\} \leq 1.$   
 (2) The definition of the conversion rate renders  $\text{volume}(\zeta_j) = \text{volume}(\zeta_{k1}) * \text{conversion}(\zeta_{k1}, \zeta_{j=k2})$ . The formula can then be obtained by iterations.

With such filtering properties, the purchase funnel (CJG) can be applied into many applications of retail business, as explained in the following section.

### CJG IN THE PM KNOWLEDGE SPACE

#### Real transaction data and its RFM model

##### Format of real transaction data

This experiment is based on the actual data of retail for big data analysis. The data was obtained from a 40-year retail brand of a company. This brand has cloud e-commerce and approximately 300 physical stores. The company is engaged in the apparel industry, and tops, jackets, shoes, bags, and accessories are top product categories. The brand has more than 40,000 active customers in the past three years.

The data of this experiment was obtained from  $N=1035056$  real transactions in 2019–2020, set  $\Xi = C \times 2^P \times 2^M \times T = [x_i, i=1, \dots, N]$ .

This data only requires four major categories, namely  $P=\{g_1=\text{tops}, g_2=\text{skirts}, g_3=\text{pants}, g_4=\text{dresses}\}$ . All the transactions are similar to that in Table 1. An actual transaction owns data attributes more than the four (customer, data, category, amount) attributed in Table 1, in which the necessary data attributes in the present study are adopted.

##### RFM model

The aforementioned transactions are real buying behaviors generated by  $N_c=36716$  active customers  $C=\{c_i, i=1, \dots, N_c\}$ . These customers can be partitioned into several groups by customer segmentation, which is a popular tool for identifying various customer groups with similar needs and behaviors [44]. Researchers divide customer segmentation into six categories, namely demographic data, geographic data, psychographic data, attitudinal data, sales data, and behavioral data, in which the visit frequency and monetary volume of sales data constitute the RFM model [45].

In practice, the RFM technique is applied with three terms: (1) recency of customer consumption/purchase (R), (2) purchase frequency over a period of time (F), and (3) purchase amount (monetary) during this period (M) [46]. Such purchase frequency (F) and purchase amount (monetary) (M) is used to classify the 36716 customers, as presented in Table 3.

**TABLE 3**  
**RFM mode of experiment data**

F/M	0- 999	10 <sup>3</sup> -(10 <sup>4</sup> - 1)	10 <sup>4</sup> -(10 <sup>5</sup> - 1)	10 <sup>5</sup> -(10 <sup>6</sup> - 1)	10 <sup>6</sup> -(4×10 <sup>6</sup> - 1)	Sum
1	532	9094	1760	0	0	11386
2-9	5	5676	13275	132	0	19088
10-99	0	0	3034	3085	4	6123
100-999	0	0	0	90	29	119
Sum	537	14770	18069	3307	33	36716

In Table 3, purchase frequency (F) is divided into four segments: F=1 for one-time customers, F = 2 ~ 9 indicates 2-9 times for repeated customers, F = 10 ~ 99 denotes 10-99 times for frequent customers, and F = 100-9999 denotes 100-999 times for very frequent customers. By contrast, purchase amount (M) is divided into five segments: 0 ~ 999, 10<sup>3</sup> ~ (10<sup>4</sup>-1), 10<sup>4</sup> ~ (10<sup>5</sup>-1), 10<sup>5</sup> ~ (10<sup>6</sup>-1), and 10<sup>6</sup> ~ (4×10<sup>6</sup>-1) corresponding to customers who consume several hundred NT\$ (new Taiwan dollars), thousands of NT\$, tens of thousands of NT\$, hundreds of thousands of NT\$, and customers of more than one million NT\$, respectively, in two years (2019–2020). Thus, 4 × 5 = 20 customer segments were possible, in which eight segments were empty.

##### Customer values and customer value segment

In Table 3, purchase frequency (F) and amount (M) are used to segment 36716 customers into 12 (=20-8) nonempty customer segments. Because customers in various segments exhibit distinct contribution to business, Table 3 is called the customer value table.

Formally, the customer value is the perception attitude of customers toward giving and getting and is related to the customer's attributes of the product and the goal or purpose that the customer wants to achieve [47, 48]. For the actual operation of customer relationship management, the customer value segments in Table 3 provides several observation points:

- (1) A total of 119 customers visit more than 100 times within two years, and the total consumption amount exceeds 100,000 NT\$, which can be the threshold of being Very Important Person (VIP).
- (2) A total of 11386 one-time customers constituted nearly 31% of total customers.
- (3) A total of 19088 repeat customers visited 2-9 times, of which 13275 customers spent 10,000 to 100,000 NT\$. These customers were the main consumers.
- (4) The spending of frequent customers who visited 9 times-99 times was evenly distributed between tens of thousands NT\$ and hundreds of thousands NT\$; however, four customers had a spending power of millions NT\$.

In the following, these customer value segments are used to analyze the knowledge space of CJ.

These customer value segments are used in the following subsections to analyze the knowledge space of CJ.

#### PMG/CJG of frequency customers

##### Transaction Data to the PMG

When PMG and CJG were investigated by the frequency customers of F=9-99 and M ≥ 1·10<sup>5</sup>, a comprehensive details of CJGs was

obtained. First, the 5056 transactions of 132 customers who visited 9-99 times and spent more than 10,000 NT\$ were obtained to conduct knowledge space analysis of CJ. As discussed, the two-year (2019-2020) transactions are categorized into four time intervals [T1,T2,T3,T4]=[2019/01/01-2019/06/30,2019/01/01-2019/12/31, 2019/01/01-2020-06-30,2019/01/01-2020/12/31] and then each PM is in the form of [g1,g2,g3,g4]. Thus, this indicates whether the customer purchased these four product categories in the time interval  $T_k$ ,  $k=1,..,4$ .

Algorithm PMG can be used to generate the corresponding PMG, as displayed in Figure 6, and Figure 6 adopts 9 PM ECs {0000, 1000, 1100,.., 1111} as its vertex set. The edges denote the directions of purchase journeys. Note that the value under purchase measure, for example, the number 4 under 1101, represents the number of elements in the PM EC.

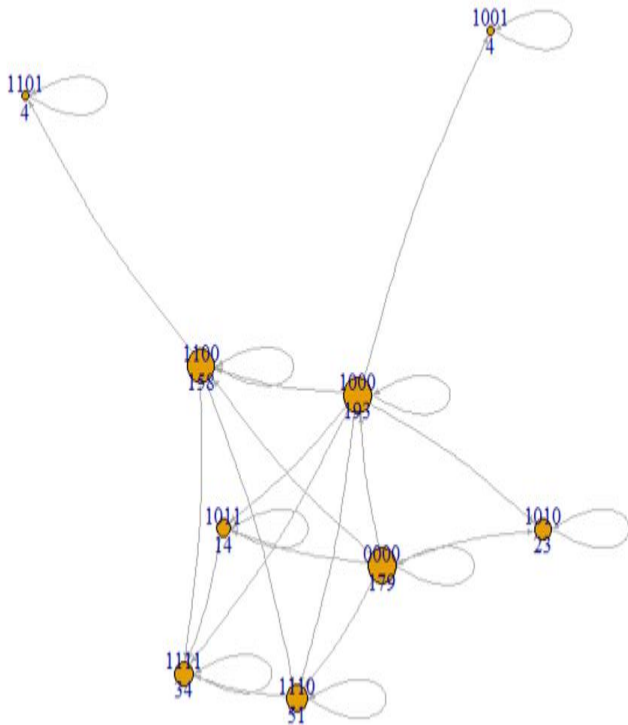


Figure 6) Purchase measure graph of 132 returning valuable customers

When self-loops are neglected, Figure 6 can form a partial ordering set, with the pseudo-source 0000 as the top node and 1001, 1010, 1101, and 1111 as the bottom nodes. To ensure such CJs obvious, the knowledge space analysis is helpful.

Knowledge space analysis and the CJG

Given Figure 6 as the PMG, Algorithm CJG can be used to generate a sophisticated CJG, as displayed in Figure 7. In a CJG, the knowledge bases are displayed in pink circles. Each knowledge state node is labeled with the CJ code, volume, and the corresponding PM. For each edge between two states, the conversion rate and customer flow are also labeled.

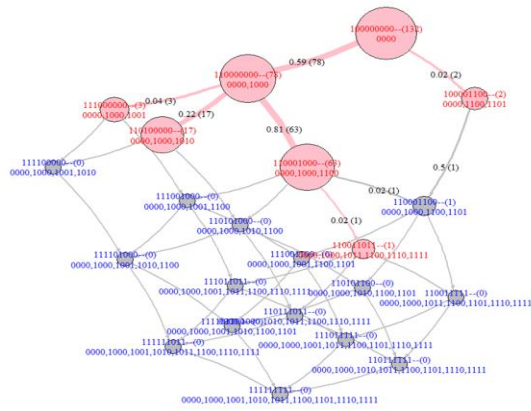


Figure 7) Customer journey graph (CJG) of 132 returning valuable customers

Data insights of Figure 7

The subgraph in Figure 7 corresponding to those pink nodes depicts the structure of knowledge bases. Two main branches of the primary knowledge bases exist: nearly 60% customers (59.1%) purchase g1 (1000) in the following choice, and only a small part of customers (1.5%) purchase g1, g3, and g4 next. The first branch (11000000) then continues to purchase g2, g3, or g4 with different probabilities (80.8%, 21.8% and 3.8%). Moreover, customers buying g1 and g2 are probably (with probability 1.6%) make more complete (g3 and g4) purchase. The knowledge base structure provides a probabilistic description of the knowledge space of the entire customer purchase journey.

Knowledge analysis of customers with different customer value segments

PMGs of different customer value segments

For the customer value table in Table 3, all the 12 customer segments in the customer value model were considered, the resultant PMGs from Algorithm PMG and their nodes are arranged in Table 4. Because the customer numbers in the last monetary segment  $10^6 \sim 4 \cdot 10^6$  are small, they are combined with the previous number  $(10^5 - (10^6 - 1))$ .

Table 4 reveals that for one-time customer and repeat customer visiting 2-9 times, the corresponding PMGs are simple ring structures, especially for those customer segments with ten thousand NT\$ ( $10^4$  -1). These simple CJs originate from the low-visit frequency and sparse purchased product categories. The largest one-category purchase is  $g_1$  (1000) and the largest two-category purchases are  $(g_1, g_2)$  and  $(g_1, g_3)$ , with 1100 and 1010, respectively.

CJGs of different customer value segments

Customers with consumption amount more than ten thousand NT\$ have more complex purchase journey, which should be explored by Customer Journey Graph (CJG, purchase funnel). For customer segments with monetary amounts greater than 10000 ( $10^4$ ), Table 5 lists all the Frequency (F), Monetary (M), customer number, customer journey codes, knowledge states, and the resultant CJGs. Each bit in a knowledge state of Table 5 corresponds to one PM in the customer journey code, for example, the second bit in knowledge states of the first data row (row (A)) denotes the 0001 PM.

**TABLE 4**  
**Various customer segments and their PMGs**

No.	F (frequency)	M (monetary)	No. of customers	purchase measure (nodes)	PMG (Purchase Measure Graph)
(A)	1	0-999	532	0000, 0010, 0100, 1000	
(B)	1	$10^3$ - ( $10^4$ -1)	9094	0000, 0001, 0010, 0011, 0100, 0101, 0110, 1000, 1001, 1010, 1011, 1100, 1110	
(C)	1	$10^4$ - ( $10^5$ -1)	1760	0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	
(D)	2-9	0-999	5	0000, 1000	
(E)	2-9	$10^3$ - ( $10^4$ -1)	5676	0000, 0001, 0010, 0011, 0100, 0101, 0110, 1000, 1001, 1010, 1011, 1100, 1101, 1110	

(F)	2-9	$10^4 - (10^5 - 1)$	13275	0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	
(G)	2-9	$10^5 - (10^6 - 1)$	132	0000, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	
(H)	10-99	$10^4 - (10^5 - 1)$	3034	0000, 0001, 0010, 0100, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	
(I)	10-99	$10^5 - (4 \times 10^6 - 1)$	3085+ 4	0000, 0010, 0100, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	
(J)	100-999	$10^5 - (4 \times 10^6 - 1)$	90+29	0000, 1000, 1010, 1100, 1101, 1110, 1111	

**TABLE 5**  
**Various Customer Segments and their CJGs**

No.	F (frequency)	M (monetary)	No. of customers	customer journey code w.r.t. purchase	(Customer journey) knowledge state	Purchase Funnel(Customer Journey Graph, CJG)
(A)	2-9	$10^4-(10^5-1)$	13275	0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	<p>1000000000000000,</p> <p>1100000000000000,</p> <p>1000010000000000,</p> <p>1000000010000000,</p> <p>1011000000000000,</p> <p>1010000010000000,</p> <p>1000100010000000,</p> <p>1001000010000000,</p> <p>1010000011000000,</p> <p>1000100011001000,</p> <p>1000101010101000,</p> <p>1000000111111111</p>	
(B)	10-99	$10^5-(10^6-1)$	132	0000, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	<p>100000000,</p> <p>110000000,</p> <p>111000000,</p> <p>110100000,</p> <p>110001000,</p> <p>100001100,</p> <p>110011011</p>	
(C)	10-99	$10^4-(10^5-1)$	3034	0000, 0001, 0010, 0100, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	<p>100000000000,</p> <p>1100100000000,</p> <p>1010100000000,</p> <p>1001100000000,</p> <p>1001100010000,</p> <p>1000111000000,</p> <p>1000101010000,</p> <p>1000101111111</p>	
(D)	10-99	$10^5-(4 \times 10^6-1)$	3085+4	0000, 1000, 0010, 0100, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	<p>10000000000,</p> <p>10010000000,</p> <p>11010000000,</p> <p>10110000000,</p> <p>10011100000,</p> <p>10011001000,</p> <p>10010101000,</p> <p>10011111111</p>	
(E)	100-999	$10^5-(4 \times 10^6-1)$	90 +29	0000,1000, 1010, 1100, 1101, 1110, 1111	<p>1000000, 1100000,</p> <p>1001000, 1110000,</p> <p>1101111</p>	



Data insights of Table 5

Several observations can be made from these purchase funnels:

- (1) As designed initially, all CJs originated from the pseudo-source (knowledge state with customer journey 0000).
- (2) For the CJGs in Table 1-5 and 2 primary knowledge bases with different customer journeys were present, in which these purchase journeys with  $g_1$  (1000) were mostly the largest volumes 10968, 78, 667+2462+6+2, 1589 and 24+94, respectively. Here,  $g_1$  in the third and fifth CJGs mix with other product categories.
- (3) CJGs not only exhibited the primary purchasing product category but the secondary or even the third product categories, shown as those knowledge states with path lengths  $\geq 2$ .
- (4) Comparing CJGs of More Valuable Customers (MVCs, with  $F=2-9$  and  $M = 10^5 \sim (10^6-1)$ ) in row (B) and more frequent customers (MFCs, with  $F=10-99$  and  $M = 10^4 \sim (10^5-1)$ ) in row (C). According to CJGs, MFCs have diversified in their primary purchasing products (110010000000, 101010000000, 100110000000 and 100011100000); whereas, MVCs ensure a more concentrated primary purchase (59.1% for  $g_1$  with 11000000 coding), and subsequent differentiate with other three secondary purchasing categories.
- (5) The last CJG (in row (E)) of the customers with the most frequent visits and most valuable amount illustrate the final mature purchase patterns, in which approximately 20% customers will purchase  $g_1$  and 80% customers purchase  $g_1$  and  $g_2$  simultaneously.

Because these CJs originate from two-year transactions, longer data may provide complete results. Commercially, three-year data are always used.

Knowledge analysis of VIPs

Customer frequency and monetary measures

In the experiments, the actual retail transaction data is used to develop PMGs and CJGs with purchase product measure  $\mu_F(\cdot)$ . In the following, the research point turns to the construction of the PMG/CJG with other purchase measures, such as purchase frequency measure  $\mu_F(\cdot)$  and/or purchase monetary (the amounts of products) measure  $\mu_M(\cdot)$ .

Here, both of these two measures are considered with (1)  $\mu_F(\zeta[\dots, T_k]) = 0, 1, 2, 3$  or 4 for the values of visiting frequency ( $\zeta[\dots, T_k]$ ) to be 0, 1, 2-9, 10-99, or 100-999, respectively; and (2)  $\mu_M(\zeta[\dots, T_k]) = 0, 1, 2, 3$  or 4 for the values of purchased amount monetary ( $\zeta[\dots, T_k]$ ) falls in 0, 0-999, 1000-9999, 10000-99999, 100000-999999, and 1000000-3999999, respectively. Next,  $\mu_{FM}(\zeta_i) = [\mu_F(\zeta_i), \mu_M(\zeta_i)] = 00 - 44$  is a combined measure of  $\mu_F(\cdot)$  and  $\mu_M(\cdot)$ , such as  $\mu_{FM}(\cdot)=23$  means frequency ( $\zeta[\dots, T_k]) = 2 - 9$  and monetary ( $\zeta[\dots, T_k]) = 10000 - 99999$ .

From the customer value table in Table 3, the customer value segments of 90 VIPs customers with  $F=100-999$  and  $M=100000-999999$  are selected in the following customer journey study.

PMG of VIPs with purchase frequencies/monetary measures

Because these 90 customers are VIPs, they generate 21545 transactions in two years and generate abundant information about purchase journey. According to the approach of this study, seven ECs (vertices), namely 00, 11, 13, 23, 33, 34, and 44, exist with each owning 93, 1, 1, 4, 8, 182, and 161 elements numbers, respectively. From the partial ordering property of CJs, the switching from one PM to another PM is used to generate the PMG, as displayed in Figure 8. By the essential nature of VIPs, the source and destination nodes of Figure 8 are 00 (the node of no consumption) and 44 (the node of  $F=100-999$  frequency and  $M=100000-999999$  consumption), respectively.

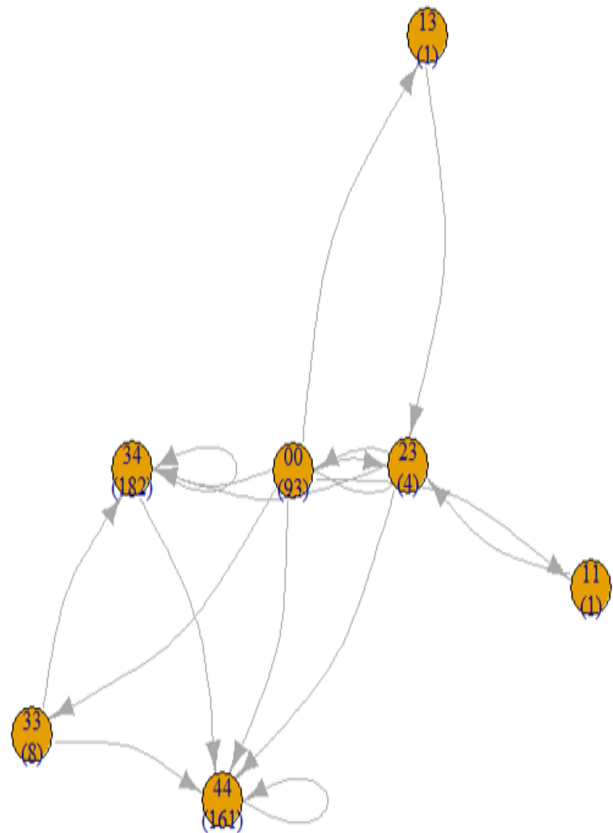


Figure 8) Purchase frequency/monetary measure graph of VIPs

CJG of VIPs with purchase frequencies/monetary measures

To systematically study CJ, Algorithm CJG was used to generate a knowledge space of CJG, as displayed in Figure 9. The knowledge bases in Figure 9 illustrate two development paths of VIPs: (1) 10 customers from pseudo-source (00) directly jump into 23, 33 (consuming 10000-99999 in 10-99 times), then 6 out of 10 customers develop into VIPs (31). (2) A small part (2) customers exhibit slow consumption 11 to 13 (once with 1000-9999 consumption) in the beginning.



Although only 10+2 customers are present in these two basic consumptions (knowledge bases), all other customers consume in mixed styles. These customers with suitable components (coordinates) can be located, as derived in Section 4.

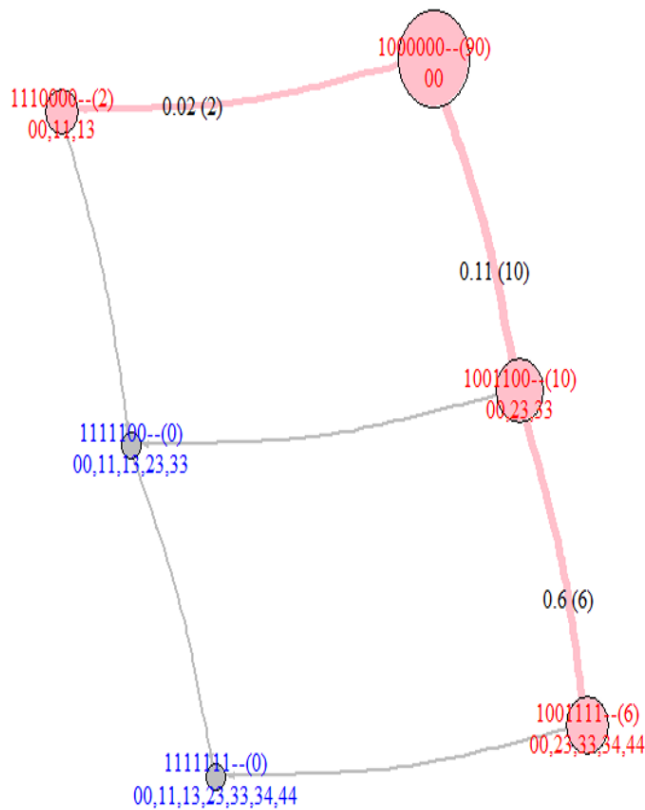


Figure 9) CJG of VIPs with purchase frequency/monetary measures

CONCLUSION

Summary

This study provides a strong theoretical foundation of CJs. Furthermore, PM graph and CJG (purchase funnel) were established to investigate the repurchase behavior in retail transactions. For repurchase, CJ and purchase funnel of retailing, Section 1 details their definitions, characteristics, and conceptual model to provide a research background. Section 2 derives the mathematical formulation of PMs and forms the PM space. Furthermore, Section 3 presented a set of purchase measure ECs and detailed construction of PMG. Section 3 removes the self-loops of a PMG, finds its terminal nodes to form partial ordering, which is used to perform the knowledge space analysis for the knowledge base of CJ. Section 4 provides quantitative values of the knowledge states of the knowledge space, such as the coordinates, the volumes, and the customer flows of state transitions. Thus, a CJG was obtained on filtering the information. This is called the purchase funnel. Furthermore, the theoretical derivation and illustrated example of purchase funnel is presented, and a complete two-year transaction data of actual retail business is presented in Section 5. Several PMGs and CJGs were established to investigate CJs effectively.

Future works

In this study, the corresponding PMGs and CJGs of various (12, here) customer segments were detailed, and the resultant knowledge bases of the primary purchases and secondary re-purchases were applied to analyze and effectively predict customer purchase behaviors. This study provides a theoretical basis for the knowledge space analysis of CJ. Future studies can focus on numerous avenues. Section 4 details the coordinates, volumes, and customer flows for obtaining CJG. Numerous knowledge states properties can be investigated in the future. The derivation of the knowledge states was obtained from Algorithms 1.28 and 1.31, whose calculation require considerable computation time and should be improved. Such applications require effective algorithms. Finally, the feasibility of these two graphs, PMGs and CJGs, should be detailed in the retail business.

FUNDING

No Funding

AUTHOR CONTRIBUTION

These authors contributed equally to this work.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

NOMENCLATURE

Special Symbols

$i, j, k$	indices
$t$	time index
$2^S$	set of the subsets of S
$v[k]$	indexed element of vector $v$
$()$	functions or mappings
$n(S)$	norm of S (the number of elements for discrete set S)
$Vset(G)$	vertex set of graph G
$Eset(G)$	edge set of graph G

English Lowercase: Elements or Vectors (bold)

English Uppercase: Sets, Vectors and Relations (often bolds)

$C = \{ c_i \}$	customer set
$P = \{ p_j \}$	product set
$M = \{ m_k \}$	monetary (payment) set
$T = [ T_k ]$	time intervals
$T$	time
$x = (c_i, \{p_j\}, \{m_j\}, t)$	transaction
$R_\mu(x_1, x_2)$	PM equivalence relation
$Q_\mu = [ q_k ] = \Xi/R_\mu$	PM equivalence class (quotient space)
$E(\Xi, \zeta, \mu)$	PM relation
$VL = [ volume(\zeta_k) ]$	number of customers in knowledge state $\zeta_k$
$FL = [ flow(\zeta_{k1}, \zeta_{k2}) ]$	number of customers switching from $\zeta_{k1}$ to $\zeta_{k2}$
$CR = [ conversion(\zeta_{k1}, \zeta_{k2}) ]$	conversion rate from $\zeta_{k1}$ to $\zeta_{k2}$
$R, F, M$	Recency, Frequencies, Monetary (consumption amount) of customers

**Greek Lowercase: measure, customer journey**

$\xi_{c_i}$	customer journey of customer $c_i$
$\zeta_{c_i, T_k}$	touchpoint of customer journey of $c_i$ at $T_k$
$\zeta = \{\zeta_i\}$	set of customer journey $\zeta_i$
$\mu(\zeta_{c_i}, T_k): \Omega \rightarrow \mathbb{R}$	purchase measure
$\delta(\mu(\zeta_{c_i}, T_k))$	purchase measure increment
$B_K = \{\beta_k\}$	PM knowledge base of knowledge space $(\zeta^*, K^*)$
	$\beta_0$ : PM origin, $\{\beta_k\}$ : PM axes, $\{\beta_{k'}\}$ : PM extended axes
$A_K = \{a_k\}$	coordinate system w. r. t. $B_K = \{\beta_k\}$
$\eta_i = \eta(\zeta_i) = [\eta_{ik}(\zeta_i)]$	customer journey code

**Greek Uppercase: space and graph**

$\Xi(C, P, M, T)$	transaction space
$\Omega(C, P, M, T) = \{\zeta_{c_i, T_k}\}$	the space of touchpoints (customer journey blocks)
$(\Xi, \zeta, \mu)$	purchase measure space
$\Psi(Q_\mu, E)$	purchase measure graph (PMG) with vertex set $Q_\mu$ and edge set $E$
$\Psi_i(Q_\mu, E)$	purchase measure graph (PMG) with vertex set $Q_\mu$ and edge set $E_i$
$(\zeta, K)$	PM knowledge structure with domain $\zeta$ and knowledge structure $K$
$(\zeta^*, K)$	quasi-ordinal knowledge structure $\zeta^*$
$(\zeta^*, K^*)$	PM knowledge space
$\Gamma(K^*, \Delta)$	customer journey graph (CJG) or called purchase funnel
$\Gamma(K^*, \Delta, VL, FL, CR)$	quantitative analysis of customer journey graph (CJG) with volume $VL$ , flow $FL$ and conversion rate $CR$ .

**REFERENCES**

1. Dodds WB, Monroe KB, Grewal D. Effects of price, brand, and store information on buyers' product evaluations. *J Mark Res.* 1991;28(3): 307-19.
2. Tsiros M, Mittal V. Regret: A model of its antecedents and consequences in consumer decision making. *J Consum Res.* 2000;26:401-17.
3. Keller KL. Building customer-based brand equity. *Mark Manag.* 2001;10:14-21.
4. Monroe KB. Pricing: Making profitable decisions McGraw Hill. 1990; 2nd ed.
5. Davidow M. Organizational responses to customer complaints. What works and what doesn't. *J Serv Res.* 2003;5(3):225-50.
6. Jones TO, Sasser WE. Why satisfied customers defect. *Harv Bus Rev.* 1995;73(6):88-99.
7. Francken DA. Postpurchase consumer evaluation, complaint actions and repurchase behavior. *J Econ Psychol.* 1983;4(3):273-90.
8. Cronin JJ, Taylor S. Measuring service quality: A re-examination and extension. *J Mark.* 1992;56(3):55-68.
9. LaBarbera PA, Mazursky D. A longitudinal assessment of consumer satisfaction/dissatisfaction: The dynamic aspect of the cognitive process. *J Mark Res.* 1983;20(4):393-404.
10. Fullerton G. The impact of brand commitment on loyalty to retail service brands. *Can J Adm Sci.* 2005;22(2):97-110.
11. Hunt KA, Keaveney SM, Lee M. Involvement, attributions, and consumer responses to rebates. *J Bus Psychol.* 1995;9(3):273-97.
12. Lovelock CH, Wright LK. Principles of service marketing and management. Prentice-Hall. 2002.
13. Kotler P. Marketing management. Prentice-Hall. 2003; 11th ed.
14. Reichheld FF, Sasser WE. Zero defections: Quality comes to services. *Harv Bus Rev.* 1990;68(5):105-11.
15. Kotler P. Marketing management: Analysis, planning, implementation and control (9th ed.). Prentice Hall Coll Inc. 1994.
16. Gronholdt L, Martensen A, Kristensen K. The relationship between customer satisfaction and loyalty: Cross-industry differences. *Total Qual Manag.* 2000;11(4-6):509-14.
17. Verhoef PC, Lemon KN, Parasuraman A, et al. Customer experience creation: Determinants, dynamics and management strategies. *J. Retail.* 2009;85(1):31-41.
18. Folstad A, Kvale K. Customer journeys: A systematic literature review. *J Serv Theory Pract.* 2018;28(2):196-227.
19. Meroni A, Sangiorgi D. Design for services. Gower. 2011. [
20. Stickdorn M, Schneider J. This is service design thinking: Basics, tools, cases. BIS Publishers. 2010.
21. Holmlid S, Evenson S. Bringing service design to service sciences, management and engineering. *Serv Sci Manag Eng Educ 21st Century.* 2008;3:341-45.
22. Patricio L, Fisk RP, Cunha JF, et al. (2011). Multilevel service design: from customer value constellation to service experience blueprinting. *J Serv Res.* 2011;14(2):180-200.
23. Zomerdijs LG, Voss CA. Service design for experience-centric services. *J Serv Res.* 2010;13(1):67-82.
24. Zomerdijs LG, Voss CA. NSD processes and practices in experiential services. *J Prod Innov Manag.* 2011;28(1):63-80.
25. Diana C, Pacenti E, Tassi R. Visualtiles: Communication tools for (service) design. In *Proc ServDes.* 2009;65-76.
26. Clatworthy S. Service innovation through touch-points: development of an innovation toolkit for the first stages of new service development. *Int J Des.* 2011;5(2):15-28.
27. Gloppen J. Service design leadership. In *Proc ServDes.* 2009;77-92.
28. Kankainen A, Vaajakallio K, Kantola V, et al. Storytelling group—a co-design method for service design. *Behav Inf Technol.* 2012;31(3):221-30.
29. Nichita ME, Vulpoi M, Toader G. Knowledge management and customer relationship management for accounting services companies. In *Proc Eur Conf Knowl Manag.* 2012;12(6):435-42.
30. Kaila S. How can businesses leverage data analytics to influence consumer purchase journey at various digital touchpoints?. *J Psychosoc Res.* 2020;15(2):699-714.
31. Whittle SR, Foster M. Customer profiling: getting into your customer's shoes. *Manag. Decis.* 1991;27(6).
32. Townsend W W. Bond salesmanship. H. Holt. 1924.
33. Lavidge RJ, Steiner GA. A model for predictive measures of advertising effectiveness. *J Mark.* 1961;25:59-62.
34. Lewis EE. Financial advertising. (The history of advertising). Levey Brothers. 1908.
35. Jansen BJ, Schuster S. Bidding on the buying funnel for sponsored search campaigns. *J Electron Commer Res.* 2011;12(1):1-18.
36. Runthecompany. Sale Funnel tools and insights. Retrieved from. Accessed June 24, 2020.

37. Lamey L, Deleersnyder B, Steenkamp JE, et al. New product success in the consumer packaged goods industry: A shopper marketing approach. *Int J Res Mark.* 2018;35(3):432-52.
38. Hamilton R, Price LL. Consumer journeys: developing consumer-based strategy. *J Acad Mark Sci.* 2019;47(2):187-91.
39. Gloppen J. Perspectives on design leadership and design thinking and how they relate to European service industries. *Des Manag J.* 2009;4(1):33-47.
40. Doignon JP, Falmagne JCl. Spaces for the assessment of knowledge. *Int J Man-Mach Stud.* 1985;23(2):175-96.
41. Doignon JP, Falmagne JC. Knowledge spaces. 1999.
42. Stahl C, Meyer D, Hockemeyer D. Package 'kst'. 2022.
43. Doignon JP, Falmagne JC. Knowledge spaces and learning spaces. Tech rep. 2015.
44. Abbasimehr H, Bahrini A. An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation. *Expert Syst Appl.* 2021;192:116373.
45. Griva A, Bardaki C, Pramatari K, et al. Retail business analytics: Customer visit segmentation using market basket data. *Expert Syst Appl.* 2018;100:1-16.
46. Hughes AM. Strategic database marketing: The masterplan for starting and managing a profitable, customer-based marketing program. Probus Pub. Co. 1994.
47. Zeithaml VA. Consumer perceptions of price, quality and value: A means-end model and synthesis of evidence. *J Mark.* 1988;52(3):2-22.
48. Woodruff RB. Customer value: The next source of competitive advantage. *J Acad Mark Sci.* 1997;25:139-53.